Yifeng Li
(Ethan Lee)

(+852) 53035114 | ethanelift@gmail.com | github.com/ylxmf2005

Education

- **Hong Kong Polytechnic University (PolyU)** — Hong Kong SAR
  Bachelor of Science (Honours) in Computer Science, CGPA: 3.404/4.3   *Sep 2023 – Jul 2027 (Expected)*

Professional Experience

- **Software Engineer Intern** — Dec 2025 – Present
  Mol (Hong Kong) Quant Limited
  - Built full-stack equity management system with contribution tracking, dividend distribution, weighted voting, and convertible bond workflows using Next.js, Supabase, and blockchain tokenization (ERC-20 on Sepolia)
  - Developed financial news aggregation platform with multi-source crawling, LLM-powered analysis (multi-provider), and vector similarity search (Milvus) using Clean Architecture (FastAPI + Celery + Redis)
  - Architected MolBot, a multi-agent financial analysis system with 5 parallel market modules (US, FX, CN, HK, Crypto) producing automated daily reports across 4 time slots, with cross-asset narrative synthesis and Xiaohongshu publishing

- **Author** — Mar 2026 – Present
  CLIClaw — Multi-Agent Orchestration Daemon (GitHub)
  - Built self-hosted agent runtime that turns Claude Code and Codex CLI into always-on autonomous agents with durable envelope routing, multi-tier memory, scheduled autonomy, and native Telegram bot integration
  - Designed pluggable multi-provider execution framework where official AI CLIs serve as native backends, inheriting tool-calling, sandboxing, and model optimizations without custom LLM wrappers
  - Implemented durable SQLite-backed envelope protocol, timezone-aware cron scheduling, team-based workspaces with shared memory, RBAC token isolation, and real-time web console (Next.js)

- **Co-author & Maintainer** — Sep 2024 – Present
  Open LLM Vtuber — Voice-Interactive AI Companion (5.3k+ GitHub Stars)
  - Co-developed cross-platform AI companion with real-time voice conversation, visual perception, Live2D avatar, and MCP-based tool calling
  - Designed multi-agent conversation system with dynamic turn-taking and hierarchical memory architecture layers for long-term context retention
  - Optimized real-time streaming pipeline through sentence-level ASR-LLM-TTS parallelization, reducing first-response latency

- **Author** — Nov 2024 – Present
  DeepLecture — AI-Native Video Learning Platform (GitHub)
  - Built AI-native video learning platform: bilingual subtitles (ASR + translation), AI dubbing with playback sync, knowledge timeline segmentation, and PDF-to-video generation with vision-based slide annotation
  - Generated learning materials (notes, Q&A, flashcards, quizzes) using multimodal grounding from video frames, subtitles, and slides

- **Co-Founder & CTO** — Feb 2025 – Nov 2025
  Navox — HKSTP Ideation Programme, PolyU Micro Fund Top Performer
  - Designed hierarchical Manager-Operator agent architecture for mobile automation with App Activity Transition Graph for cross-screen navigation and parameterized Shortcut Library for workflow reuse
  - Implemented 3-tier execution system (UI-Tars, OmniParser, coordinate memory) with intelligent fallback handling for exceptions (e.g. popups)

Research Experience

- **Research Assistant** — Sep 2024 – Present
  PolyU URIS Program, Natural Language Processing Group @PolyU — *Supervised by Prof. Wenjie Li*
  - Investigating text style transfer methods for LLM training data augmentation
  - Built full-stack simulation sandbox (Unity + FastAPI + Vue3) for evaluating LLM social reasoning capabilities

- **Research Internship** — Oct 2023 – Jun 2024
  Tencent Security Xuanwu Lab, Tencent Spark Program
  - Co-authored SecEval (100+ stars), the first benchmark for evaluating LLM cybersecurity knowledge, featuring 2,000+ questions across 9 domains from authoritative sources (MIT/Berkeley textbooks, OWASP, MITRE ATT&CK)
  - Contributed to data engineering for EDR Alert Analysis Robot (95.3% expert consistency, 80K alerts/day); co-invented patent CN120821638A
  - Conducted adversarial research on LLM security vulnerabilities including prompt injection, jailbreak, and leakage attacks with corresponding defense methods

Awards & Achievements

- **AdventureX 2025 Hackathon — AI Watch Buddy** — Jul 2025
  Team Captain, Best Voice Innovation Award by Tenframework — *Hangzhou, China*
  - Built proactive AI video companion with real-time visual understanding, streaming action system and expressive Live2D avatar integration
  - Designed context engineering pipeline for video-grounded conversation with structured action schema and interrupt-aware dialogue management
  - Won Best Voice Innovation Award among 800+ participants selected from 5,000+ global applicants in 120-hour hackathon

- **ICPC Asia Regional Contest 2023** — 2023
  Bronze Medal

- **National Olympiad in Informatics in Provinces (NOIP)** — 2021
  First Prize

Core Competencies

AI/Agent Systems	Agentic Loop / Workflow, Tool Use, MCP, A2A, Agentic RAG, Prompt / Context Engineering, Multi-Agent Orchestration, Vector Database
Voice AI Pipeline	VAD (Voice Activity Detection), Turn Detection, TTFT (Time to First Token) Optimization, ASR/LLM/TTS Streaming, Multimodal Grounding
Tech Stack	Python, C++, TypeScript · Full-stack Web (React, Next.js, FastAPI, Supabase) · Databases (SQL, NoSQL, Vector) · ML/LLM Frameworks (PyTorch, LangChain) · CI/CD & Containers (GitHub Actions, Docker) · AI-assisted Dev (Claude Code, Codex)